Detrending, denoising, and prediction of geophysical processes

Jianbo Gao
PMB Intelligence LLC, West Lafayette, IN 47906
jbgao@pmbintelligence.com, gao@ece.ufl.edu

Wen-wen Tung
Department of Earth & Atmospheric Sciences,
Purdue University, West Lafayette, IN 47907
wwtung@purdue.edu

Jing Hu,
PMB Intelligence LLC & Affymetrix, Santa Clara, CA
jhu@pmbintelligence.com

Outline

- Backgrounds
- Local smooth adaptive filtering (LSAF) for
  - detrending
  - denoising
- Detecting chaos from noisy time series by SDLE
  - low- and high-dimensional chaos
  - intermittent chaos
- Quantifying predictability by pseudo-ensemble approach
- Example application: sunspot & river runoff dynamics
- Future perspectives

Backgrounds and motivational questions

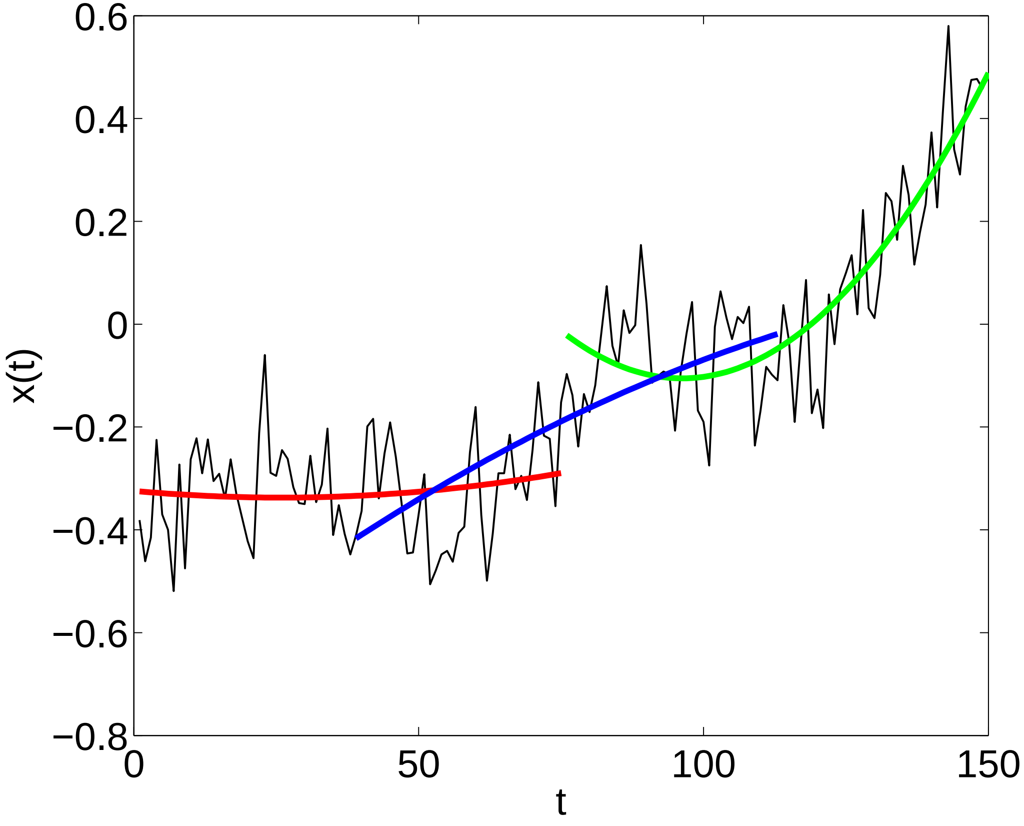
- Predictability of dynamical systems is an issue of enormous theoretical and practical importance in many disciplines of science and engineering.
- The issue is especially challenging in geosciences, since
  - Geophysical processes often contain complicated trends
  - Measurement data typically are very noisy
  - Often a mathematical model is not available for performing ensemble forecasting
- **Question**: Can we quantify the intrinsic predictability of a dynamical process using only a scalar time series?
- **Challenges**: need effective detrending/denoising algorithms to clean up the data, find the best model for the process, then do prediction

Detrending based on empirical mode decomposition

- Wu, Huang, et. al. (PNAS 2007):
  systematic, iterative method; among the best
- Basic idea:
  - Find local maxima, connect them by a cubic spline line; do the same for the local minima, and take the mean of the two envelopes
  - Subtract the mean from the data, and repeat the procedure for the remaining data
- Problems:
  - Stoppage criterion is somewhat subjective
  - The decomposition is not unique and is sensitive to noise
  - Cannot deal with chaotic signals
  - Interpretation is difficult for random fractal signals

Detrending using local methods: Challenges

- Segmentation and fitting causes jumps and discontinuities around the boundaries, thus severely distort the high-frequency components of the original signal



Local smooth adaptive filtering (LSAF): algorithm

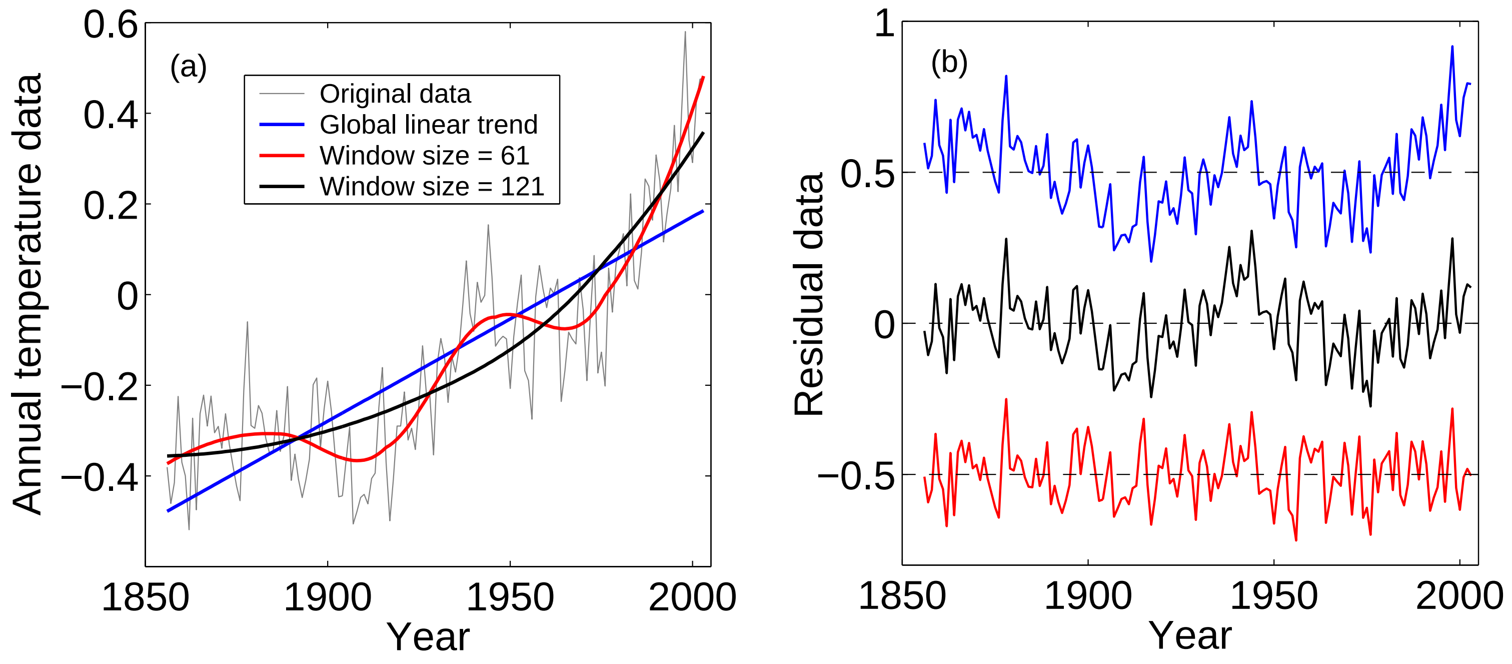
- Partition a time series into segments (or windows) of length $2n+1$ points, where neighboring segments overlap by $n+1$ points
- For each segment, fit a best polynomial of order K
- Denote the fitted polynomial for the i-th and $(i+1)$-th segments by $y^{(i)}(l_1),\ y^{(i+1)}(l_2),\ l_1, l_2 = 1, \cdots, 2n+1$, respectively.
- Define the trend for the overlapped region as

$$y^{(c)}(l) = \left(1 - \frac{l-1}{n}\right) y^{(i)}(l+n) + \frac{l-1}{n} y^{(i+1)}(l), \quad l = 1, 2, \cdots, n+1$$

- The trend is smooth at the non-boundary points, and has at least the right- or left-derivative at the boundary points
- The parameters may be determined by requiring that the variance of the detrended data no longer decreases significantly when K is further increased and/or $2n+1$ is further decreased

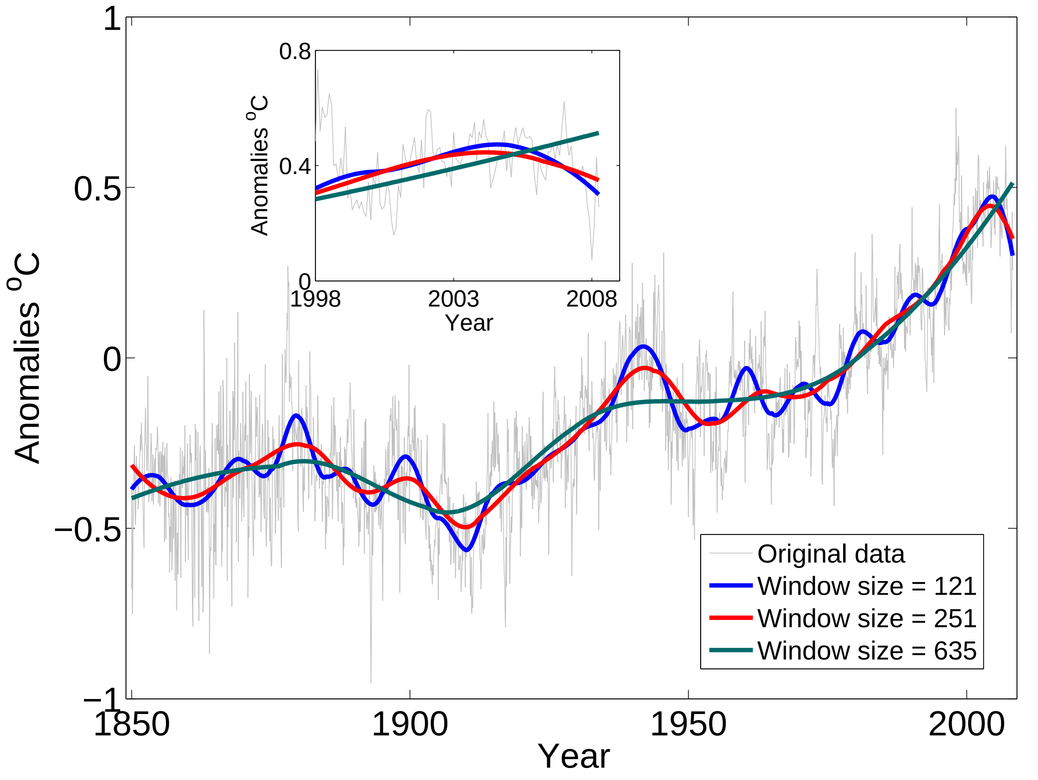
Determining trends in the yearly GSTA data

- With these window sizes, our results are very similar to those of Wu et al;
- By adjusting window sizes, residual noise can be made even smaller

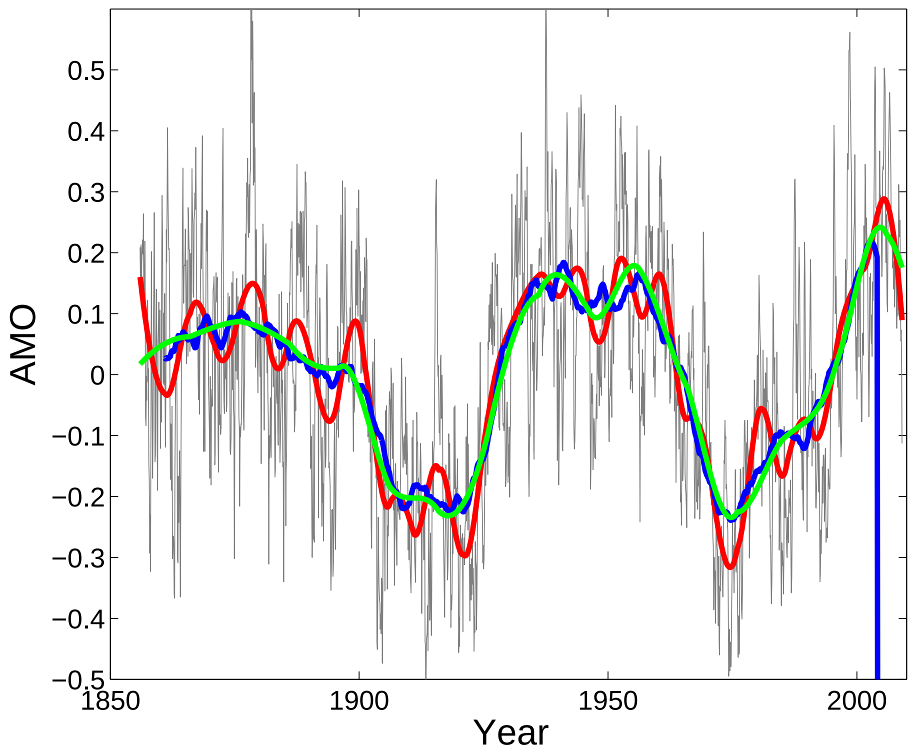


Determining trends in the monthly GSTA data

- The global warming trend is obvious
- There also exist local cooling spans



Atlantic Multidecadal Oscillation (AMO)
(http://www.cdc.noaa.gov/data/climateindices/List)



- AMO: calculated from the Kalplan SST (black & blue: unsmoothed and smoothed (window size 121 months) data, Enfield et al. GRL 2001)
- Green & red: adaptive trends with window sizes 121 and 241 months, respectively; smoother than the blue curve

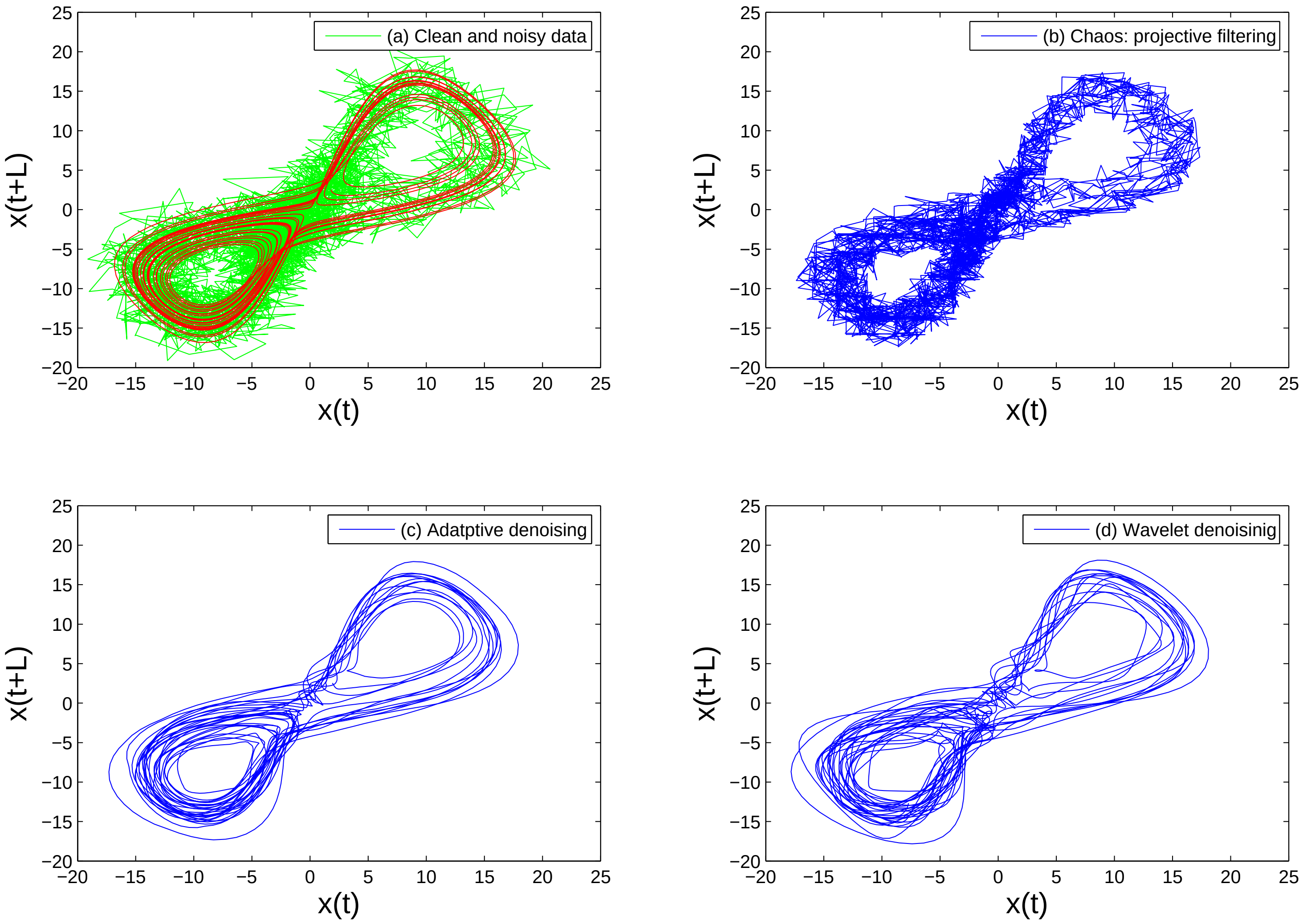
Denoising

- Chaotic Lorenz '63 model:

$$\begin{aligned}\dot{x} &= -10(x-y)+D_1\eta_1(t),\\ \dot{y} &= -xz+28x-y+D_2\eta_2(t),\\ \dot{z} &= xy-\tfrac{8}{3}z+D_3\eta_3(t),\end{aligned} \tag{1}$$

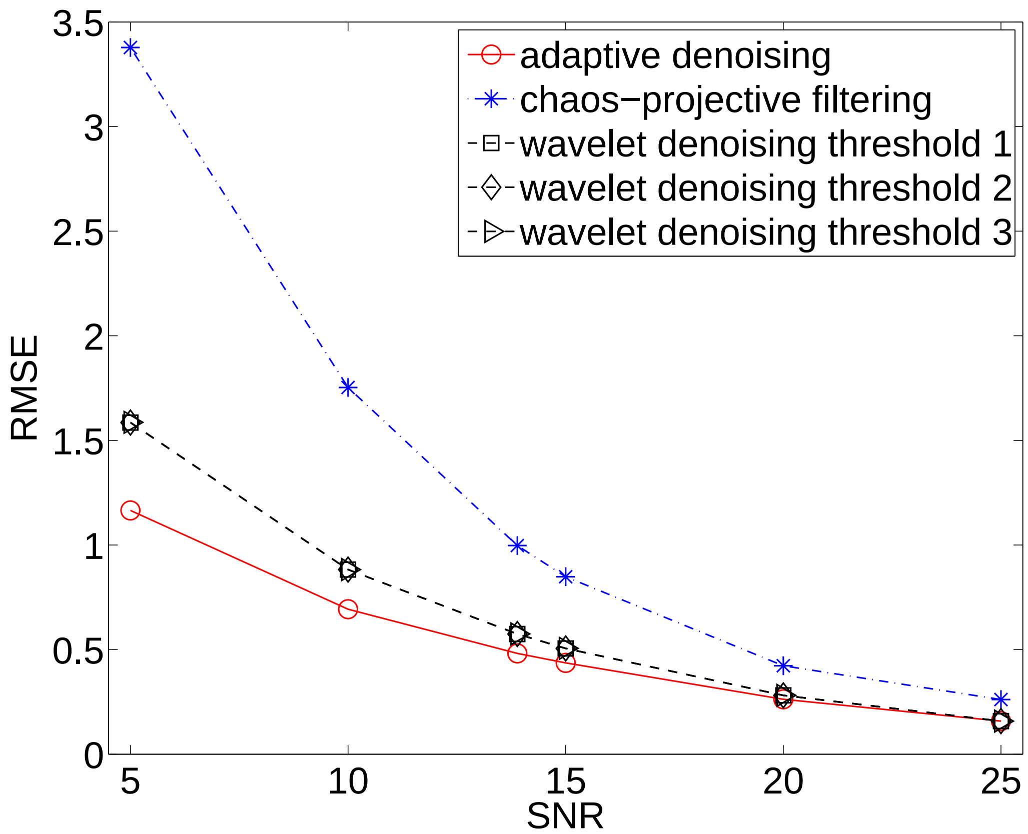
- Measurement noise: $x(t)+n(t)$; $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[x(i)-\hat{x}(i)]^2}$
- Dynamical noise: noise is in the equations of the system ($D_i \neq 0$); *RMSE* cannot be defined; effectiveness of denoising can be evaluated through recovery of chaotic signatures
- Experimental data: both measurement and dynamical noise may exist; *RMSE* cannot be defined

LSAF of chaotic Lorenz data: SNR = 13.89 dB



LSAF of chaotic Lorenz data: performance

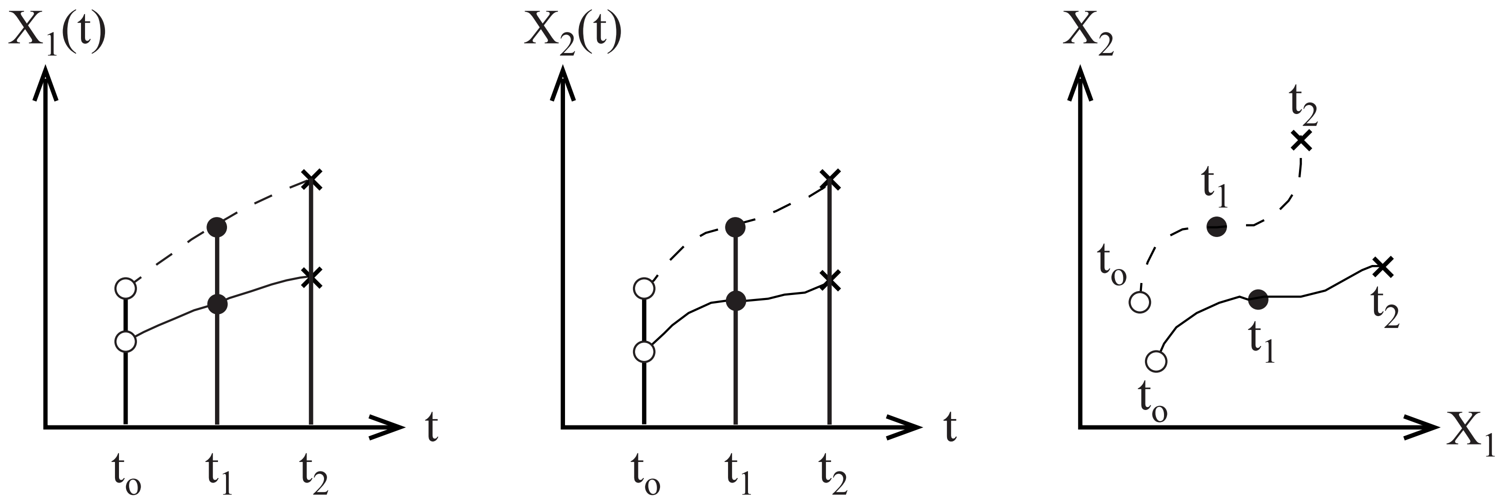
- LSAF can reduce both measurement and dynamical noise (the latter follows Cauchy distribution); more effective than linear filters as well as chaos and wavelet based approaches (and therefore, is the most effective)



Introduction to chaotic dynamics: The concept of phase space & transformation

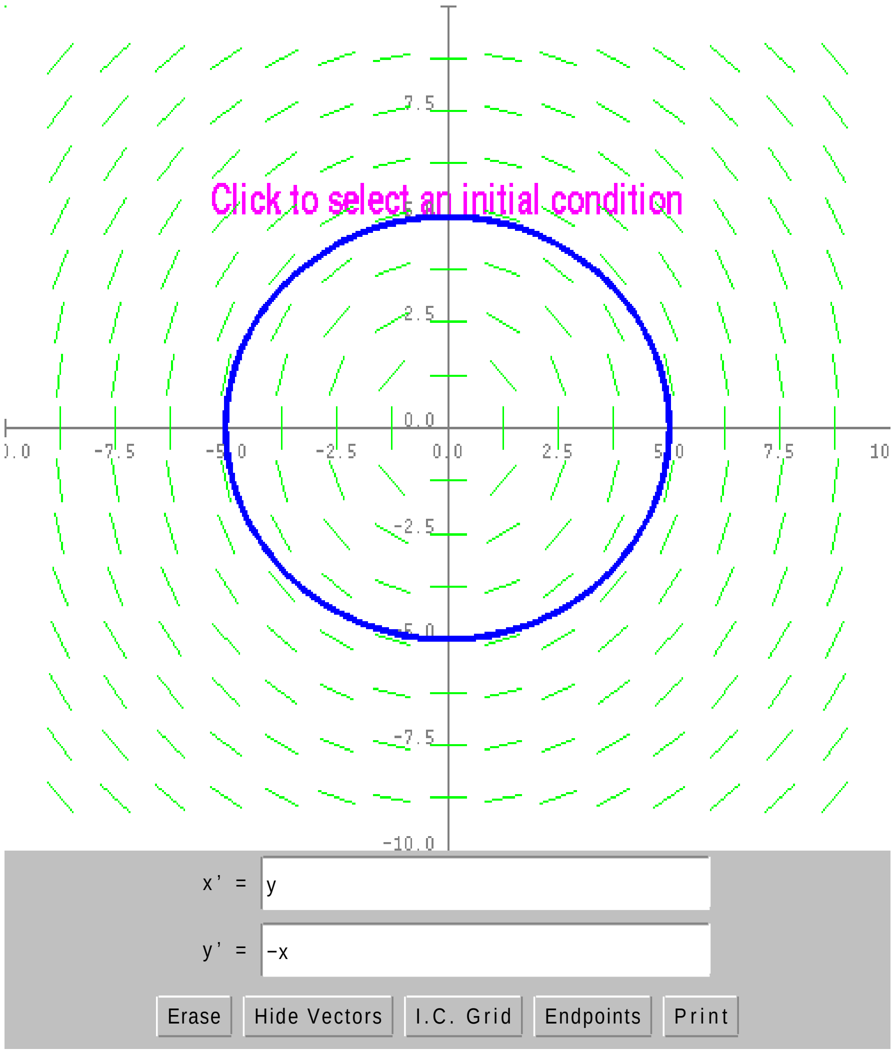
$$\frac{dX_1}{dt} = f_1(X_1, X_2), \qquad \frac{dX_2}{dt} = f_2(X_1, X_2)$$

$$(X_1(0), X_2(0)) \longmapsto (X_1(t), X_2(t)) = T^{(t)}(X_1(0), X_2(0))$$



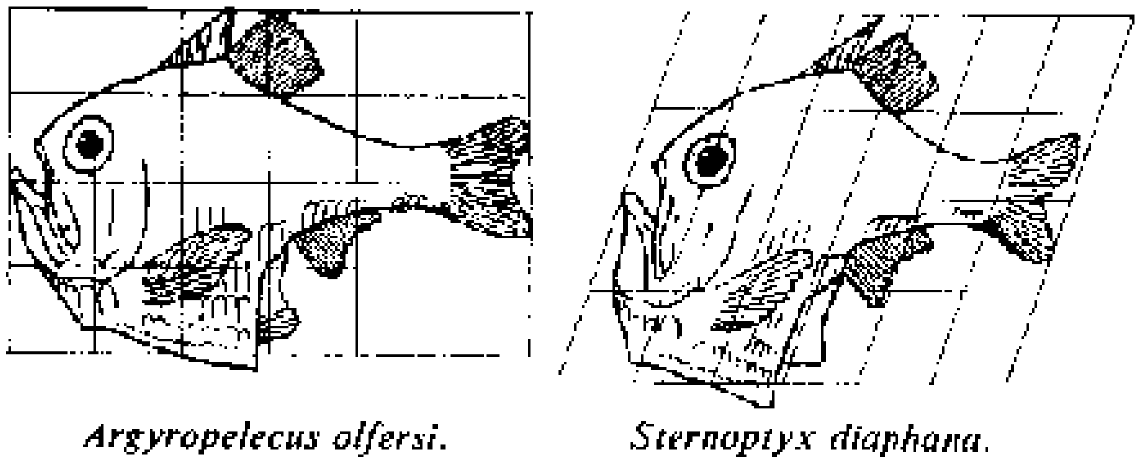
An example: phase plane of the equation $\ddot{x} = -x$

(Rewrite as: $\dot{x} = y$, $\dot{y} = -x$)



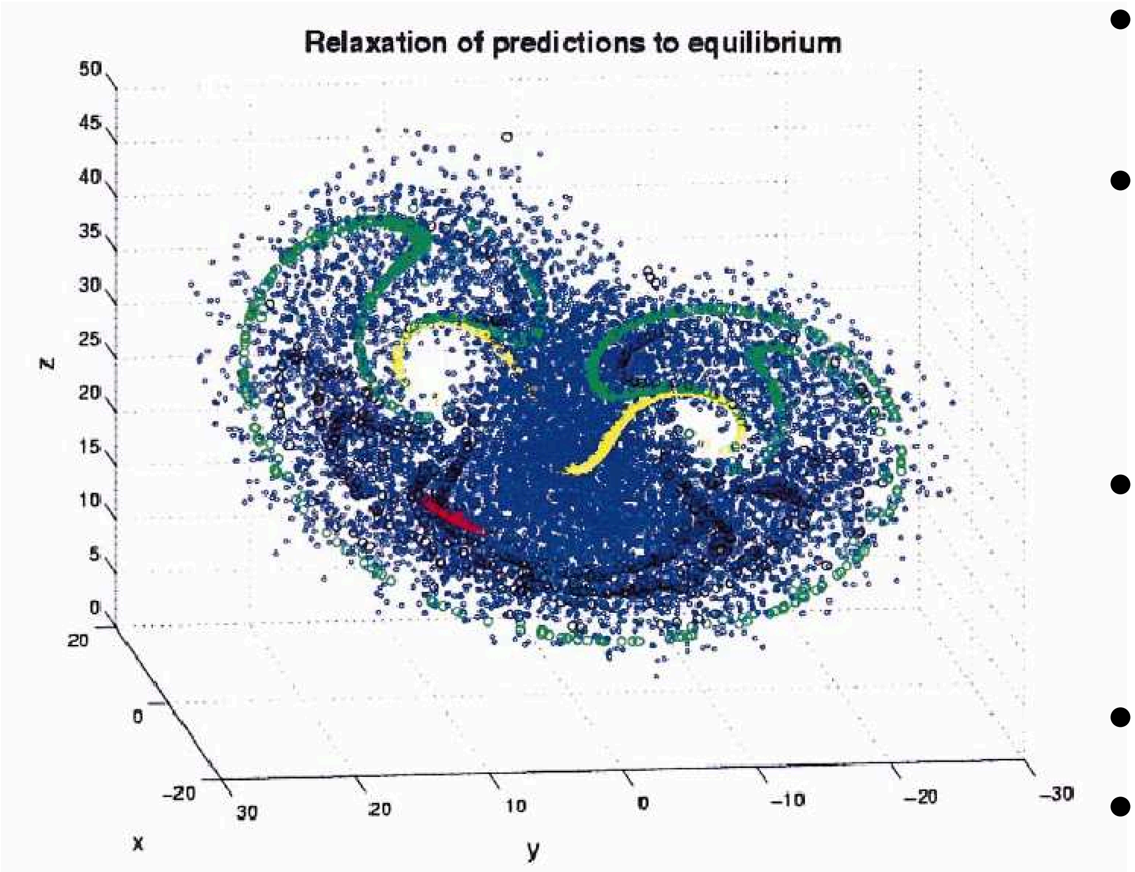
The action of flow: Transformation

- Example: Sir D'Arcy Thompson's "Growth and form"— **Fish transformation** (simple plane transformations can bring two different types of fish together: different fishes have the same origin)



- **Chaotic** transformation: the head & tail of a fish get mixed up
- In cartoon pictures, face gets badly twisted
- Dust swept off the sky

Ensemble forecasting



Kleeman (2002): Different colors show the ensemble behavior at different times

- Choose a number of initial conditions
- Form neighborhoods of small sizes around the chosen initial conditions as ensembles
- Monitor the evolution of the ensembles with time
- Assess information loss
- **Critical observation: Information loss depends on the size of ensembles**

Dynamical predictability: Major research efforts

- Dynamical approach: construct suitable forecast ensembles by
  - generating Lyapunov and bred vectors (Toth and Kalnay, 1992, 1993, 1997, Kalnay 2003)
  - generating singular vectors (Palmer et al. 1993, Reynolds and Palmer 1998),
  - sampling the fastest growing directions of the phase space (Ehrendorfer and Tribbia 1997)
  - examining the dependence of prediction efficiency on ensemble size (Buizza and Palmer 1998)
- Statistical approach: quantify the prediction utility by
  - Shannon entropy (Carnevale and Holloway 1982, Schneider and Griffies 1999, Leung and North 1990)
  - **relative entropy** (Roulston and Smith 2002, Kleeman 2002; Kleeman et al. 2002; Majda et al. 2002; Cai et al. 2004; Kleeman and Majda 2005; Abramov et al. 2005, Haven et al. 2005)

Dynamical predictability: Motivational questions

- For chaotic systems, is the decay rate of relative entropy related to the Kolmogorov entropy?
- What are the connections between
  - different information theoretic approaches?
  - dynamical and statistical approaches?
- Currently, with simple models, one would choose as many ensembles as possible, with each ensemble containing a large number of members
- When the forecast models become increasingly complicated, however, one would only be able to afford a small number of ensembles, each with limited number of members, thus sacrificing estimation accuracy of the forecast errors
- From a single copy of scalar dataset or trajectory, can we get information similar to what can be obtained from a large number of ensembles?
- Positive answers imply enormous reductions in computational complexities and data storage and tremendous improvements in the estimation accuracy of forecasts

Quantification of information flow in a dynamical system

- Entropy characterizes the rate of creation of information in a system
- Covering a phase space with boxes of size ε, where $\varepsilon \to 0$. Let p_i be the probability that box i is visited by the trajectory and I_0 be the initial entropy. We have (Atmanspacher and Scheingraber 1987)

$$I(\varepsilon, t) = -\sum p_i \ln p_i = I_0 + Kt,$$

  where K is the Kolmogorov-Sinai (K-S) entropy
- For deterministic chaotic systems, due to exponential divergence, the number of phase space regions available to the system after a time T is $N \propto e^{(\Sigma\lambda^+)T}$, where λ^+ are positive Lyapunov exponents. Therefore, $I(T) = -\sum_{i=1}^{N} p_i(T) \ln p_i(T) = (\sum \lambda^+)T$, or $K = \sum \lambda^+$
- $K = 0$ and ∞ for deterministic regular & random systems

The decay of the relative entropy in ensemble forecasting

- Relative entropy $$R = \sum_i p_i \ln\left(\frac{p_i}{q_i}\right)$$

  $\{p_i\}$: probabilities associated with the forecast ensembles; $\{q_i\}$: equilibrium or climatological distribution of the ergodic system (also called the natural measure)

- We have proven that when the ensembles are chosen to represent the natural measure, and that the initial period for the ensembles to evolve onto the **most unstable** direction is short enough, then
$$R(t) = I_{max} - \int_0^t h(\varepsilon_t)dt$$
  - Same Equation holds for the decay of the Shannon entropy

- For chaotic systems, for a considerable range of finite scale ε, $h(\varepsilon_t) \approx const$, therefore, $R(t) = I_{max} - Kt$

How to compute $\int_0^t h(\varepsilon_t)dt$

- Need to know functional form of $h(\varepsilon_t)$
- Also need to know how ε_t grows with t
- Rich theoretical results for $h(\varepsilon_t)$ exist (Gaspard & Wang 1993)
  - For deterministic chaos, when $\varepsilon_t \to 0$, $h(\varepsilon_t) \sim K$
  - For independent noise, $h(\varepsilon_t) \sim -D_I \ln \varepsilon_t$, where D_I is the information dimension
  - For $1/f^{2H+1}$ processes, $h(\varepsilon_t) \sim \varepsilon_t^{-1/H}$, where $0 < H < 1$ is the Hurst parameter — depending on whether H is smaller than, equal to, or larger than 1/2, the process is said to have anti-persistent, short-range, or persistent long-range correlations

    — $H = 1/3$ for Kolmogorov's $-5/3$ turbulence energy spectrum
- Numerically, it is difficult to calculate $h(\varepsilon_t)$ accurately
- Can infer $h(\varepsilon_t)$ from the scale-dependent Lyapunov exponent (SDLE) $\lambda(\varepsilon_t)$; also can obtain ε_t from $\lambda(\varepsilon_t)$

Characterizing chaos by scale-dependent Lyapunov exponent (SDLE)

(Gao et al. Phys. Rev. E, 2006; Wiley book, 2007, Chaos, 2009)

- Consider an ensemble of trajectories in **phase space**
- Denote the initial separation between two nearby trajectories by ε_0, and their **average separation** at time t and $t+\Delta t$ by ε_t and $\varepsilon_{t+\Delta t}$, respectively
- Being defined in an average sense, ε_t and $\varepsilon_{t+\Delta t}$ can be readily computed from any processes, even if they are non-differentiable
- When $\Delta t \to 0$, SDLE $\lambda(\varepsilon_t)$ is defined by

$$\varepsilon_{t+\Delta t} = \varepsilon_t e^{\lambda(\varepsilon_t)\Delta t} \qquad \text{or} \qquad \lambda(\varepsilon_t) = \frac{\ln \varepsilon_{t+\Delta t} - \ln \varepsilon_t}{\Delta t}$$

- Equivalently, we have a differential equation for ε_t,

$$\frac{d\varepsilon_t}{dt} = \lambda(\varepsilon_t)\varepsilon_t$$

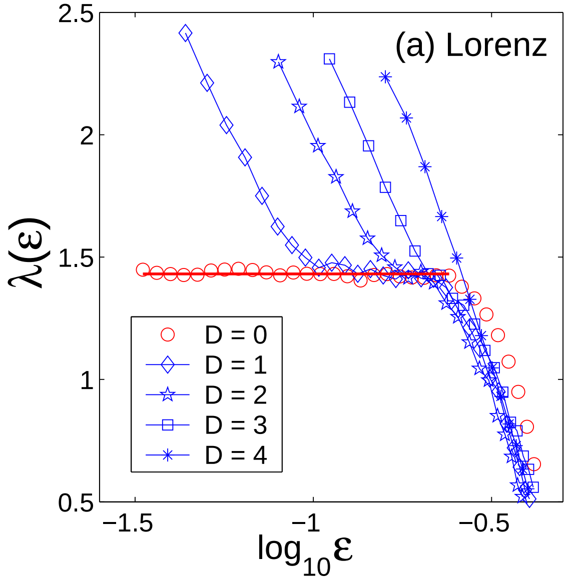
SDLE $\lambda(\varepsilon)$ for chaos, noisy chaos, & noise-induced chaos

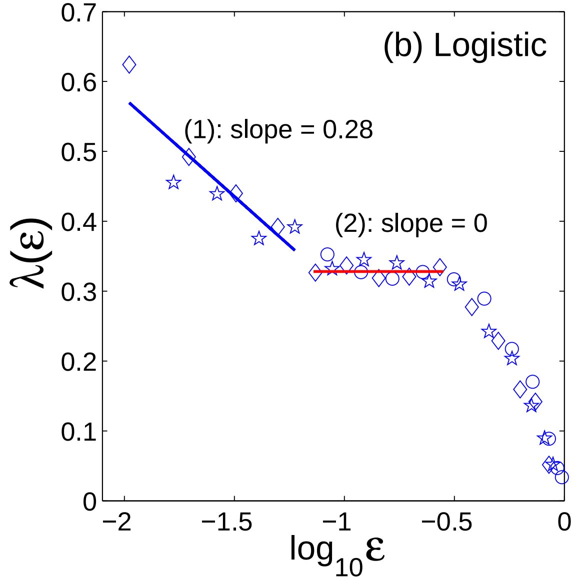
- Chaos: $\lambda(\varepsilon) \approx$ const (largest positive Lyapunov exponent)
- Noisy chaos & noise-induced chaos: $\lambda(\varepsilon) \sim -\gamma \ln \varepsilon$ on small scales
- (i) Stochastic Lorenz ('63) system

  (ii) Noisy logistic map

  $x_{n+1} = \mu x_n(1 - x_n) + P_n, 0 < x_n < 1, \mu = 3.74, \sigma_{P_n} = 0.002$

  —without noise, motion is periodic — Noise-induced chaos

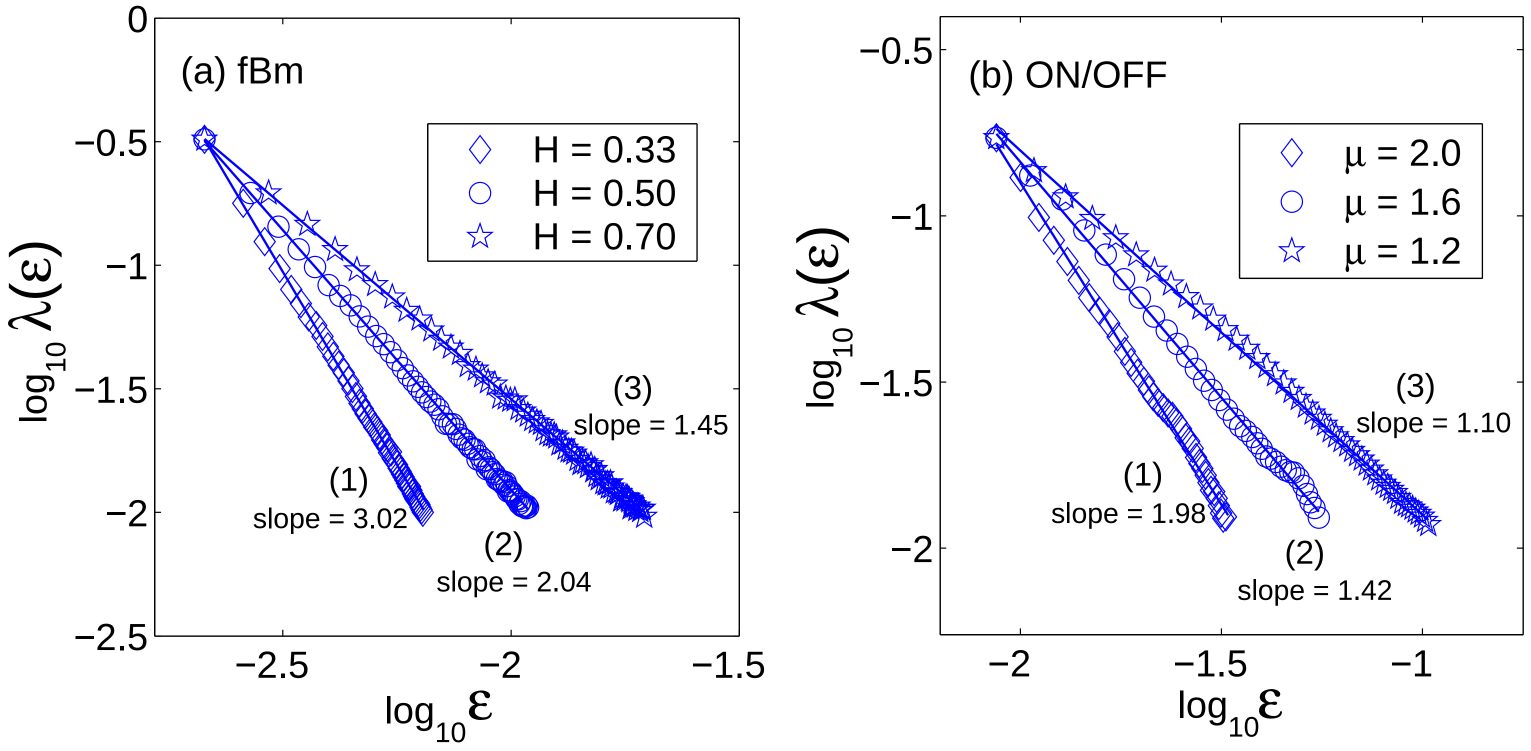




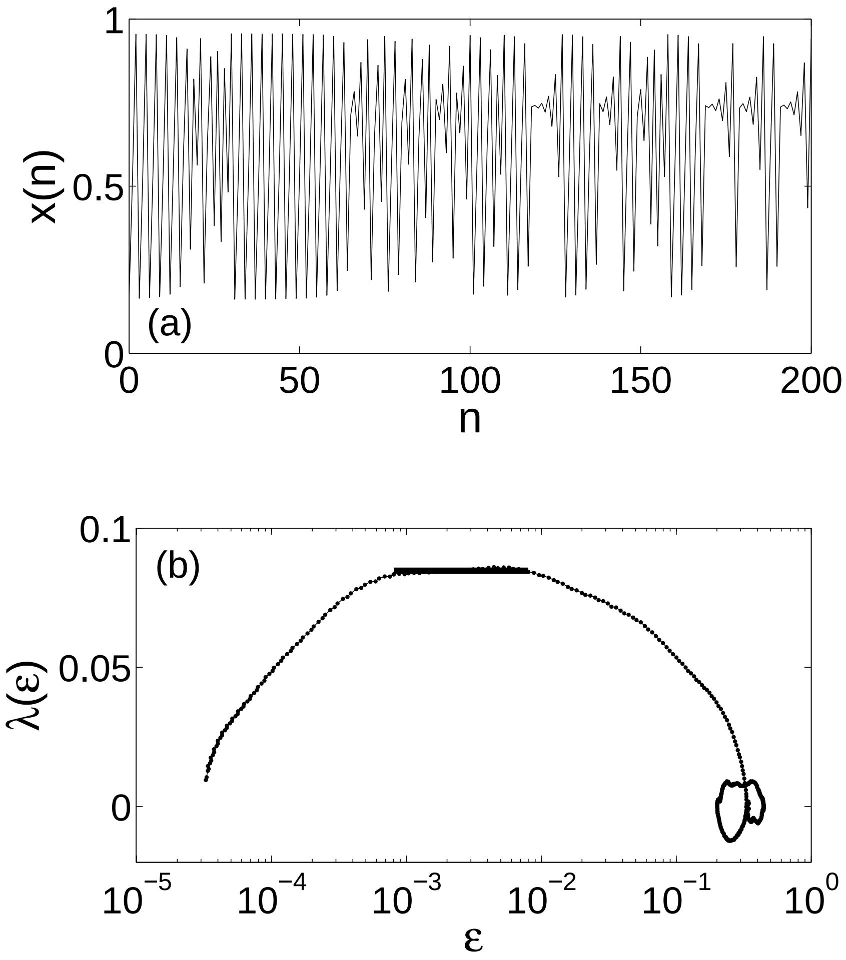
Power-law scaling of $\lambda(\varepsilon)$ for $1/f^{2H+1}$ processes

Can prove $\lambda(\varepsilon) \sim \varepsilon^{-1/H}$

For ON/OFF intermittency, $H = (3-\mu)/2$



Detecting intermittent chaos



- Intermittent chaos: regular and chaotic motions co-exist; chaotic phase can be much shorter than regular phase
- Very difficult to characterize
- Existing methods cannot detect such motions from noisy time series
- SDLE easily works, due to scale separation property
- Example: logistic map $x_{n+1} = ax_n(1-x_n), \ a = 3.8284$

Physical significance of the SDLE

- $1/\lambda(\varepsilon)$ amounts to the error doubling time
  — larger doubling time means longer prediction time scale
- The first estimate of the doubling time was 5 days, given by the Mintz-Arakawa two-layer model (Charney et al. 1966)
- With greater computational power and model complexity, one would expect doubling time to increase; however, the estimate of the doubling time has been decreasing
- A recent estimate (Simmons & Hollingsworth, 2002): < 2 days
- Lorenz suggests that the major factor for the decrease of the doubling time has been the decrease in spatial resolution
- $\lambda(\varepsilon) \sim -\ln\varepsilon$ and $\lambda(\varepsilon) \sim \varepsilon^{-1/H}$ are more relevant to reality

The pseudo-ensemble technique

Essence: ensemble forecasting equivalent based on a time series

- Define a sequence of "shells" indexed as k:

$$\varepsilon_k \leq \|V_i - V_j\| \leq \varepsilon_k + \Delta\varepsilon_k,\ |i-j| > w,$$

  where V_i, V_j are vectors sampled from a single trajectory, or vectors reconstructed from a time series $x_1, x_2, \cdots$ using Taken embedding theorem,

$$V_i = [x_i, x_{i+L}, \cdots, x_{i+(m-1)L}],$$

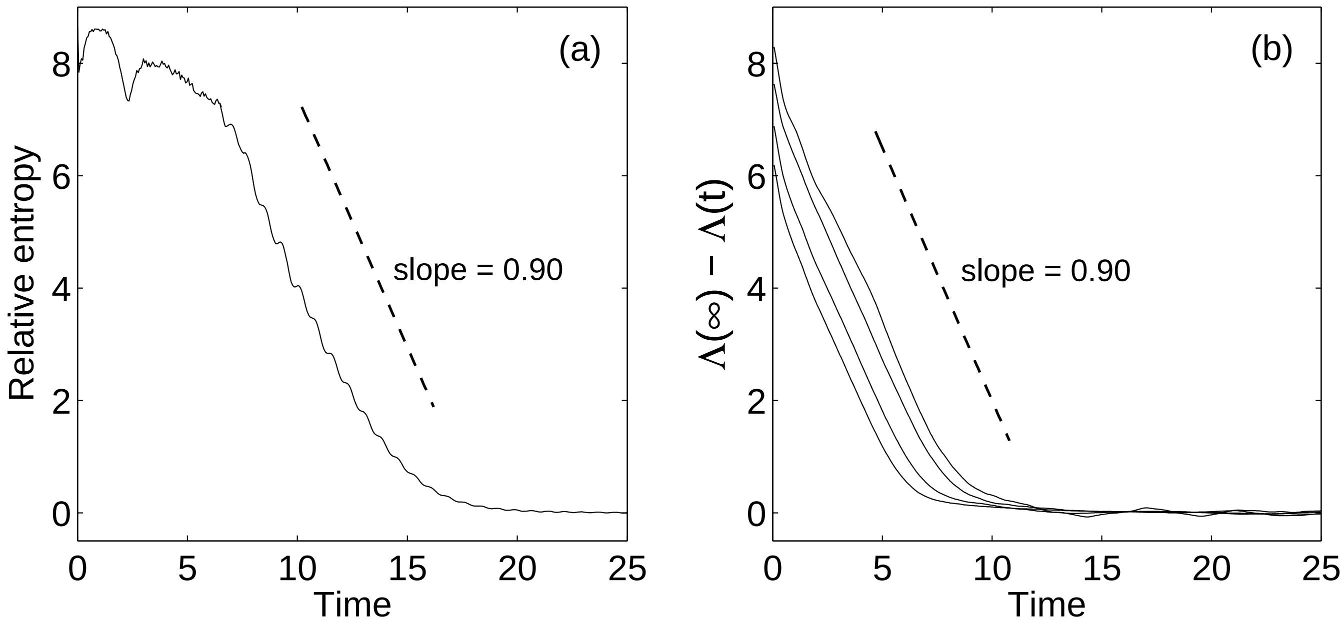
  where m and L are embedding dimension and delay time, respectively

- Computation of the SDLE (Gao et al. 2006)
  $\lambda(\varepsilon_t) = \Big\langle \ln\|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| - \ln\|V_{i+t} - V_{j+t}\| \Big\rangle / \Delta t,$
  where the angle brackets denote average within a shell

- Can prove
  $\int_0^t \lambda(\varepsilon_t)dt = \Lambda(t) = \Big\langle \ln\|V_{i+t} - V_{j+t}\| - \ln\|V_i - V_j\| \Big\rangle = \ln\varepsilon_t - \ln\varepsilon_0$
  $\Lambda(t)$: Time-dependent exponent (TDE) curves (Gao & Zheng, 93,94)

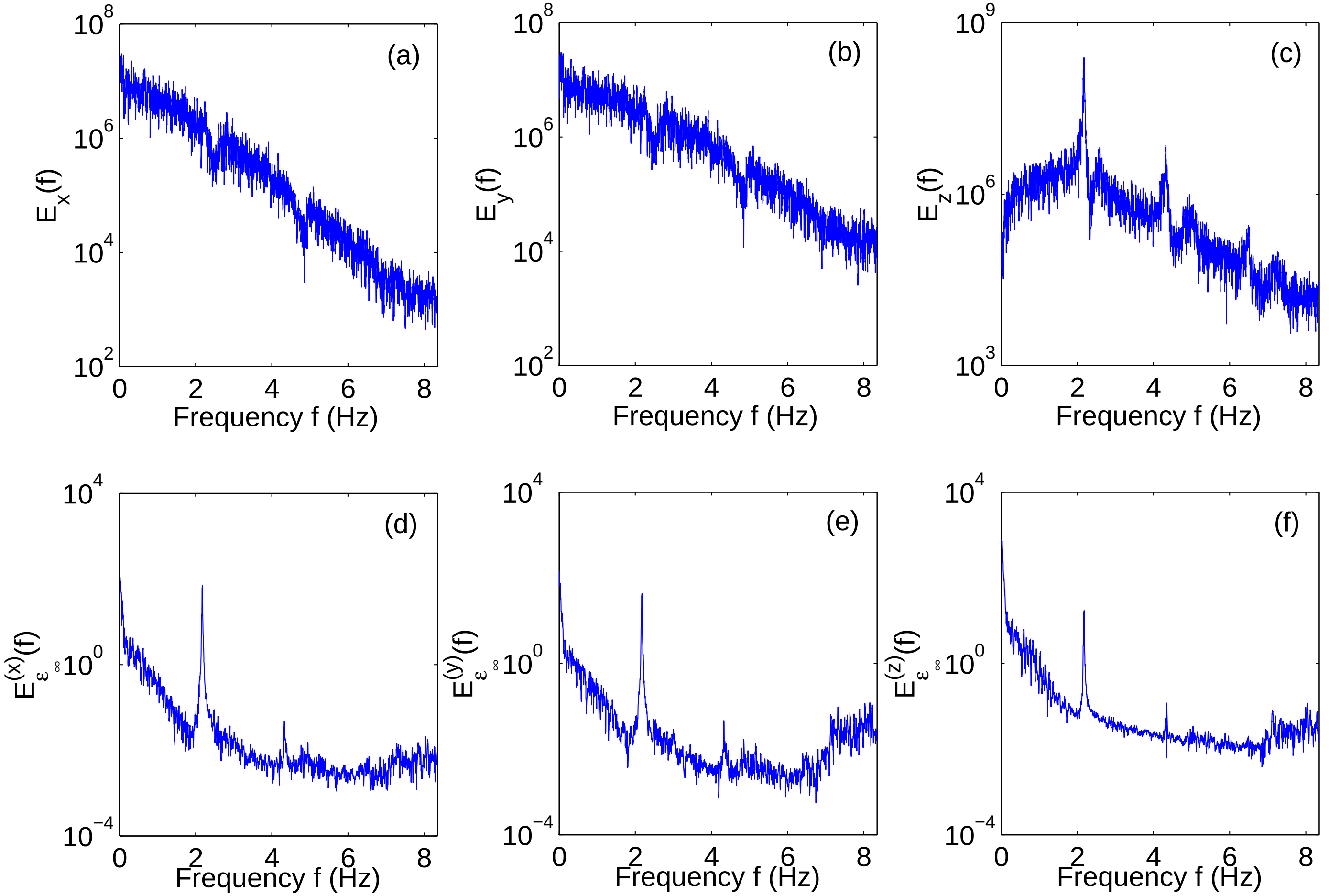
Prediction in the Lorenz '63 system

- Direct calculation of the relative entropy: partitioned the attractor into $512 \times 512 \times 512 = 2^{27}$ boxes; used 10^8 points sampled at $\tau = 0.06$; constructed 100 ensembles each having 10^6 members
- 4 TDE curves were computed using an ergodic solution of 10^5 points
- Fig.(b) takes a few minutes while (a) takes $\sim$ 100 hours to compute
- **Reduction in computational complexity and data storage by more than 4 orders of magnitude**



Predicting large scale motions in the Lorenz ('63) system

Characteristic scale ε_t: $\lambda(\varepsilon_t) \approx 0$

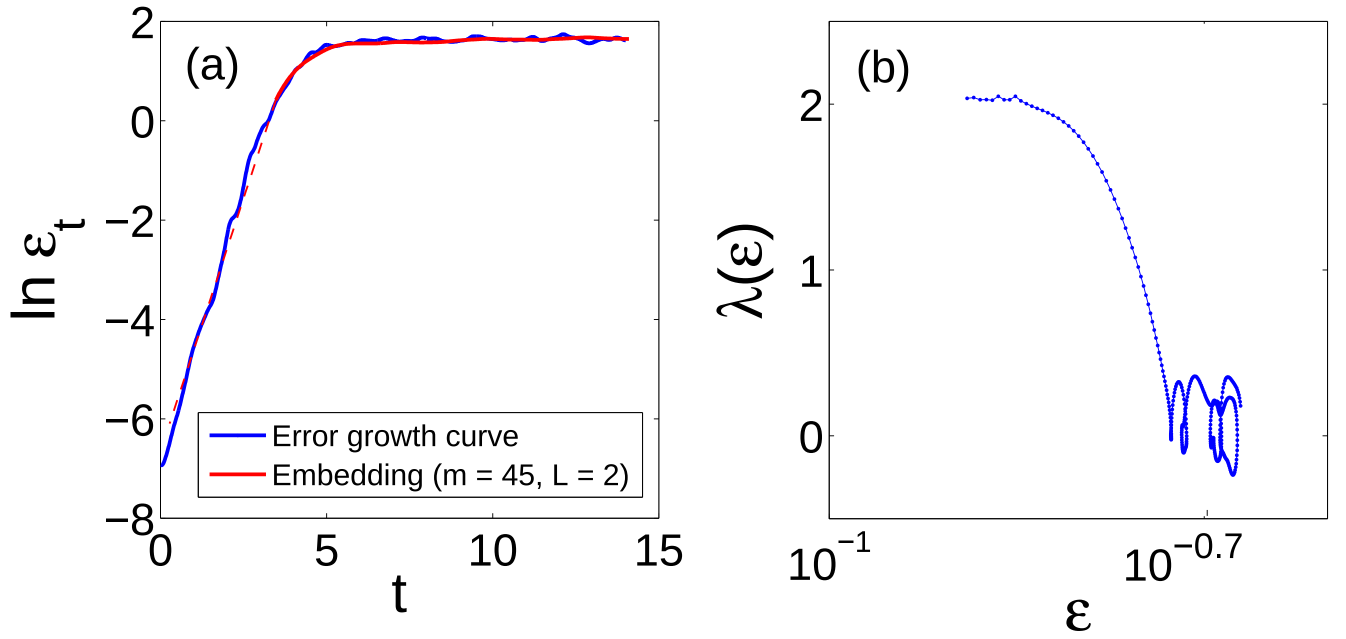


Prediction in the Lorenz '96 model

- The model is supposed to represent a 1-D atmosphere; F is a positive constant, t is (non-dimensional) time, and X_n are values for some scalar atmospheric quantity on N equally spaced latitude circle

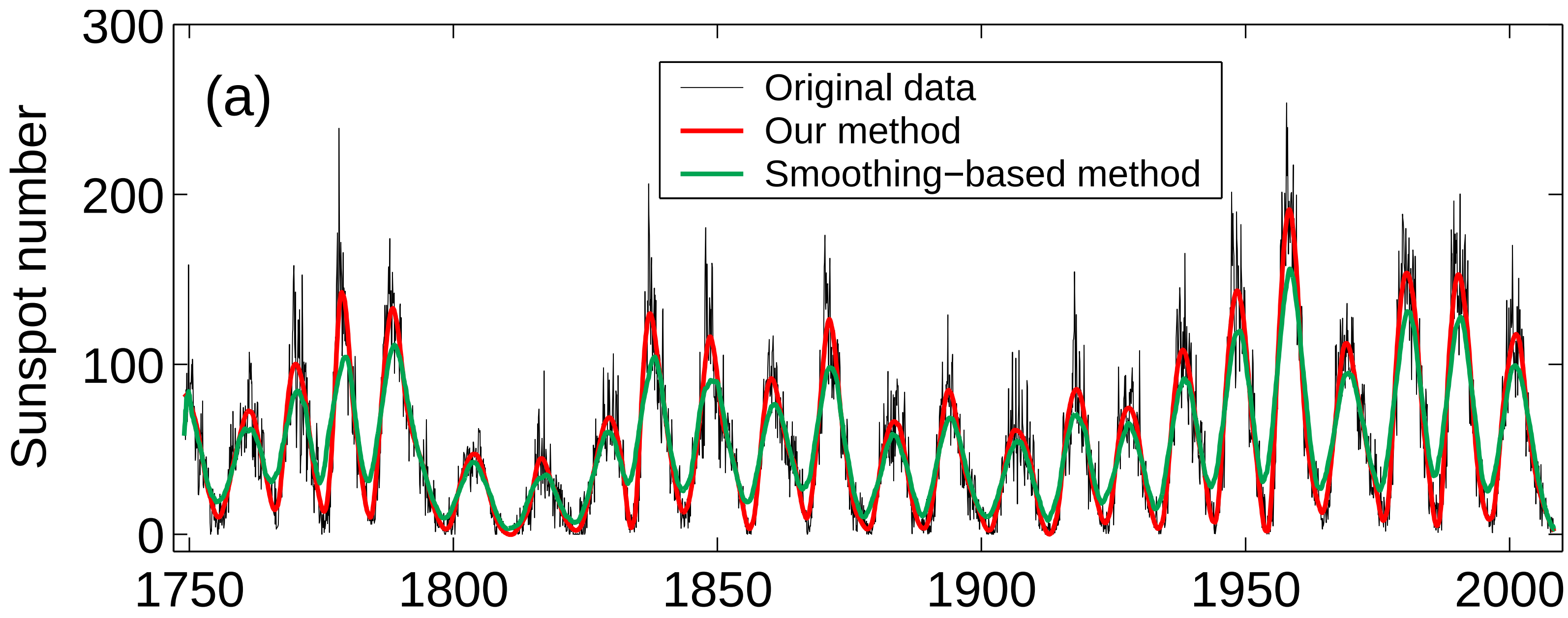
$$dX_n/dt = -X_{n-2}X_{n-1} + X_{n-1}X_{n+1} - X_n + F, \quad n = 1, 2, \cdots, N$$

- $N = 40, F = 8$ are chosen here; there are 13 positive Lyapunov exponents, $D \approx 27.1$
- Can extrapolate to small scales to recover information not resolved by a single dataset

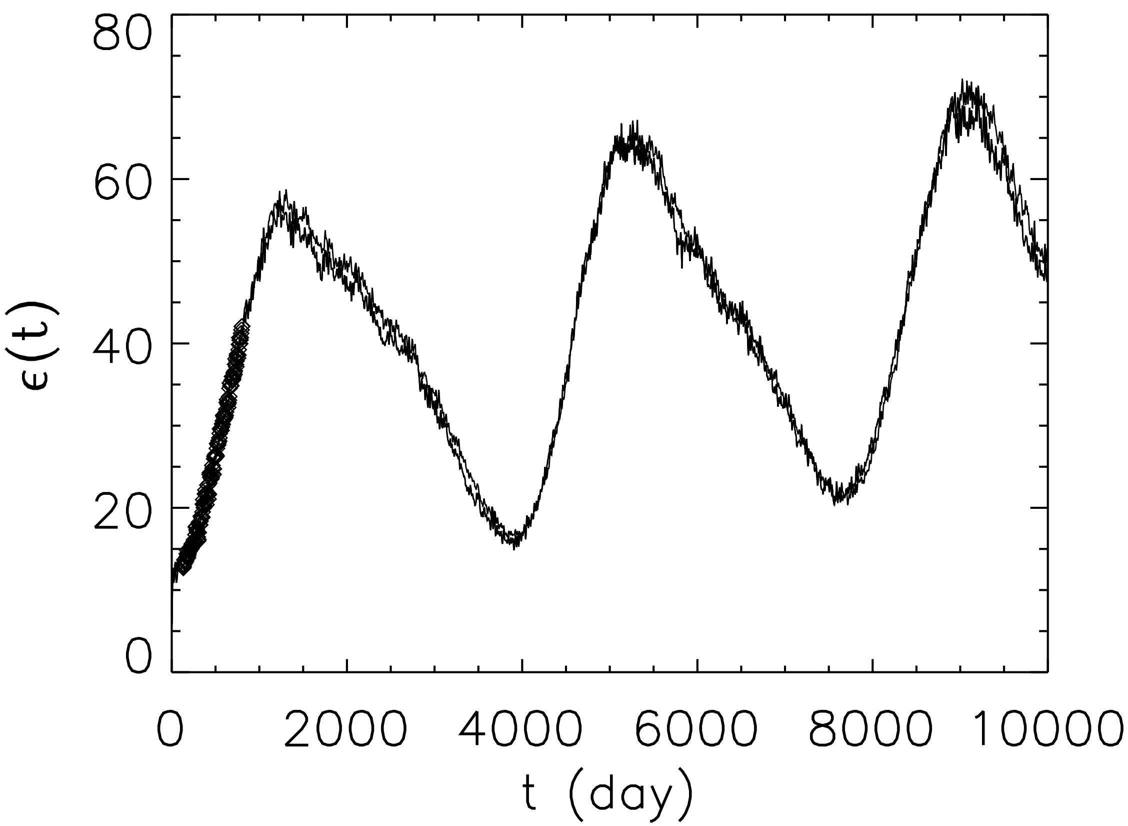


Dynamical origin of the sunspot number 11 year cycle

- After the 11 year cycle is filtered out, FFT can no longer reveals the 11 year cycle from the detrended data
- FFT of SDLE's characteristic scale of the detrended data still can
  – **Sunspot minima and maxima dynamically are different**

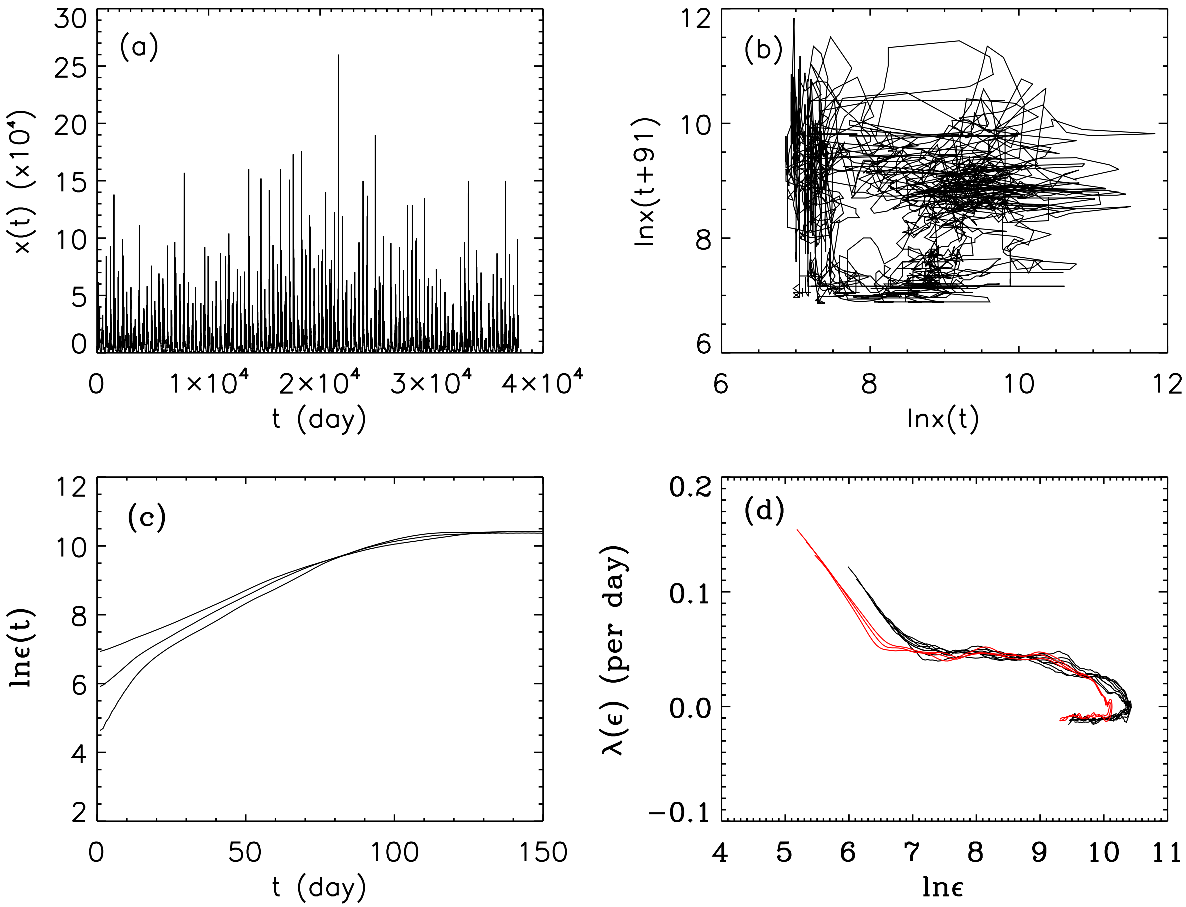


Predictability of sunspot numbers



Example application: Umpqua river runoff dynamics

- First time definitive evidence of chaos in river runoff dynamics
- The error growth curves set the limit for prediction



Quantifying predictability in medium- and high-dimensional dynamical systems

- Infinite-dimensional noisy dynamical systems: on small scales $\lambda(\varepsilon) \sim -\ln\varepsilon$; it can be resolved by around 10^4 points
- Infinite-dimensional $1/f^{2H+1}$ processes: $\lambda(\varepsilon) \sim \varepsilon^{-1/H}$; it can also be resolved by around 10^4 points
- Other infinite-dimensional stochastic processes (including Levy processes, stochastic oscillations): behavior of $\lambda(\varepsilon)$ is well-defined when there are about 10^4 points
- Medium-dimensional dynamical systems (such as Lorenz '96 model): **Here arises a major challenge!** — a given dataset defines ε_{min}. What's the behavior of $\lambda(\varepsilon)$ when $\varepsilon < \varepsilon_{min}$?
  — **Extrapolate the scaling laws for $\lambda(\varepsilon)$ to those scales!**
- Spatiotemporal systems: work with the coefficient time series of EOFs

Parameterization of error growth curves

- For ergodic systems (chaos, noisy chaos, & stochastic systems)
  - If $\lambda(\varepsilon_t) = \lambda_1$, then $\varepsilon_t = \varepsilon_0 e^{\lambda_1 t}$
  - If $\lambda(\varepsilon_t) = -\gamma \ln \varepsilon$, then $\ln \varepsilon = \ln \varepsilon_0 e^{-\gamma(t-t_0)}$, $\gamma = -D \ln \varepsilon_0 / I(\varepsilon_0)$ where ε_0 is the scale at t_0, D is the fractal dimension, and $I(\varepsilon_0)$ is the initial amount of information
- For non-ergodic systems ($1/f^{2H+1}$ & Levy processes)
  - For $1/f^{2H+1}$, $\lambda(\varepsilon) \sim \varepsilon^{-1/H}$, $\varepsilon \sim t^H$
  - Similarly for Levy processes
- For multiscale systems, there may exist different scale ranges where $\lambda(\varepsilon_t)$ has well-defined behavior; use $\varepsilon(t)$ for each scale range correspondingly
- Lorenz (1982): $\frac{d\varepsilon}{dt} = \alpha\varepsilon - \beta\varepsilon^2$, which amounts to $\lambda(\varepsilon) = \alpha - \beta\varepsilon$
  Simmons & Hollingsworth (2002): $\frac{d\varepsilon}{dt} = \gamma + \alpha\varepsilon - \beta\varepsilon^2$, which amounts to $\lambda(\varepsilon) = \frac{\gamma}{\varepsilon} + \alpha - \beta\varepsilon$

Model error and predictability

- Mode reduction is a hot topic
- Example: Lorenz '96 model
  - 1-layer model for a 1-D atmosphere; F is a positive constant, t is (non-dimensional) time, and X_n are values for some scalar atmospheric quantity on N equally spaced latitude circle

$$dX_n/dt = -X_{n-2}X_{n-1} + X_{n-1}X_{n+1} - X_n + F, \quad n = 1, 2, \cdots, N$$

  - 2-layer model:

$$dX_n/dt = -X_{n-2}X_{n-1} + X_{n-1}X_{n+1} - X_n + F - (hc/b)\sum_{j=1}^{J} Y_{j,n}$$

$$dY_{j,n}/dt = -cbY_{j+1,n}(Y_{j+2,n} - Y_{j-1,n}) - cY_{j,n} + (hc/b)X_n$$

  - Stochastic 1-layer model
- SDLE can be used to quantify the differences among them

Summary

- We have developed a versatile adaptive algorithm for detrending and denoising, which is more effective than existing methods
- We have shown that SDLE can characterize all known models of complex time series, and even detect high-dimensional and intermittent chaos
- The pseudo-ensemble approach can effectively quantify the predictability of a dynamical system using only time series data
- Application domains are vast, including geophysical sciences and biomedical engineering
- For more backgrounds, see
  Gao et al., 2007: **Multiscale Analysis of Complex Time Series — Integration of Chaos and Random Fractal Theory, and Beyond**, Wiley Interscience

Acknowledgments

- This work is partially supported by NSF grants CMMI-0825311 and CMMI-0826119
- Mr. Hussain Sultan and Ms. Lei Yang of Univ of Florida contributed to the denoising work; Drs. Yinhe Cao and Jing Ai contributed to part of the prediction work

- Thanks for your time and interests
- Looking for collaborations